Cell

Molecular Architecture of the Mos1 Paired-End Complex: The Structural Basis of DNA Transposition in a Eukaryote

Julia M. Richardson,[1,*] Sean D. Colloms,[2] David J. Finnegan,[1] and Malcolm D. Walkinshaw[1]
[1]School of Biological Sciences, University of Edinburgh, The King's Buildings, Mayfield Road, Edinburgh, EH9 3JR, Scotland
[2]Xermit, 44 Colchester Drive, Glasgow, G12 0NF, Scotland
*Correspondence: jrichard@staffmail.ed.ac.uk
DOI 10.1016/j.cell.2009.07.012

SUMMARY

A key step in cut-and-paste DNA transposition is the pairing of transposon ends before the element is excised and inserted at a new site in its host genome. Crystallographic analyses of the paired-end complex (PEC) formed from precleaved transposon ends and the transposase of the eukaryotic element Mos1 reveals two parallel ends bound to a dimeric enzyme. The complex has a *trans* arrangement, with each transposon end recognized by the DNA binding region of one transposase monomer and by the active site of the other monomer. Two additional DNA duplexes in the crystal indicate likely binding sites for flanking DNA. Biochemical data provide support for a model of the target capture complex and identify Arg186 to be critical for target binding. Mixing experiments indicate that a transposase dimer initiates first-strand cleavage and suggest a pathway for PEC formation.

INTRODUCTION

Transposition of mobile DNA elements—sequences that can jump from one place to another in the genome—has strongly influenced evolution and is a current source of genome instability and diversity (Biemont and Vieira, 2006). As revealed by sequencing efforts, transposable elements and their remnants make up a large proportion of typical eukaryotic genomes (for example, ~45% of the human genome). Previously considered as parasitic or junk DNA, it is now emerging that transposable elements have been recruited by their hosts to provide new cellular functions. For example, the V(D)J recombination system, which generates antibody diversity, is believed to have evolved from a eukaryotic transposon (Kapitonov and Jurka, 2005), and transposable elements can also play an important role in epigenetic regulation (Slotkin and Martienssen, 2007). The ability of transposons to integrate into the genome is being exploited to develop vectors for insertional mutagenesis (Dupuy et al., 2005), genome manipulation (Robert and Bessereau, 2007), transgenesis (Robinson et al., 2004) and gene therapy (Ivics and Izsvak, 2006).

One class of transposable elements, DNA transposons, move from one genomic location to another by a cut and paste mechanism, mediated by transposon-encoded transposase proteins. Transposases are sequence-specific nucleases and strand transferases that catalyze transposition through an ordered series of events: (1) sequence-specific binding of transposase to the terminal inverted repeats (IR) present at each end of the transposon, (2) pairing of the transposon IRs in a paired-end complex (PEC), (3) cleavage of one or both DNA strands at each transposon end, (4) capture of target DNA, and (5) strand transfer to insert the transposon at a new site. This is illustrated in Figure 1A for the eukaryotic *mariner* transposon *Mos1* of *Drosophila mauritiana*.

The Mos1 transposase catalytic domain contains a metal-binding DDD/E motif within a catalytic RNaseH-like structural scaffold (Richardson et al., 2006; Rice and Baker, 2001) that is conserved in other eukaryotic transposases (e.g., Hermes), RAG1 recombinase, prokaryotic transposases (e.g., Tn5 and Mu) and retroviral integrases (e.g., HIV-1). The triad of catalytic residues in the DDD/E motif forms binding sites for two divalent metal ions (Lovell et al., 2002; Richardson et al., 2006), that may enable catalysis of strand cleavage and strand transfer to proceed by a two metal-ion mechanism (Figure 1B) (Yang et al., 2006). Excision of elements from the Tc1/*mariner* family of eukaryotic transposons proceeds by two hydrolysis reactions that are usually staggered in their positions along the DNA (van Luenen et al., 1994; Luo et al., 1998; Lampe et al., 1996). Mos1 transposase cuts the transposon IR sequences with a 3 bp stagger: 3 nucleotides inside the transposon at the 5′ ends and exactly at the junction of IR and flanking DNA at the 3′ ends (Dawson and Finnegan, 2003). The two cleavage events are sequential: the nontransferred strands (NTS) are cleaved to release transposon 5′ ends before cleavage of the transferred strands (TS) to release transposon 3′ ends. While first-strand cleavage can occur without pairing of the ends, second strand cleavage requires PEC formation (Dawson and Finnegan, 2003). The excised transposon is integrated at a new site by transfer of the transposon 3′OH ends to staggered positions in the target DNA (Figure 1B). Tc1/*mariner* elements insert with a 2 bp stagger specifically into TA di-nucleotide target sites, and repair of the resulting single strand gaps leads to duplication of the target TA.

While the core catalytic domain is conserved among all members of the DDE super-family of transposases, the details

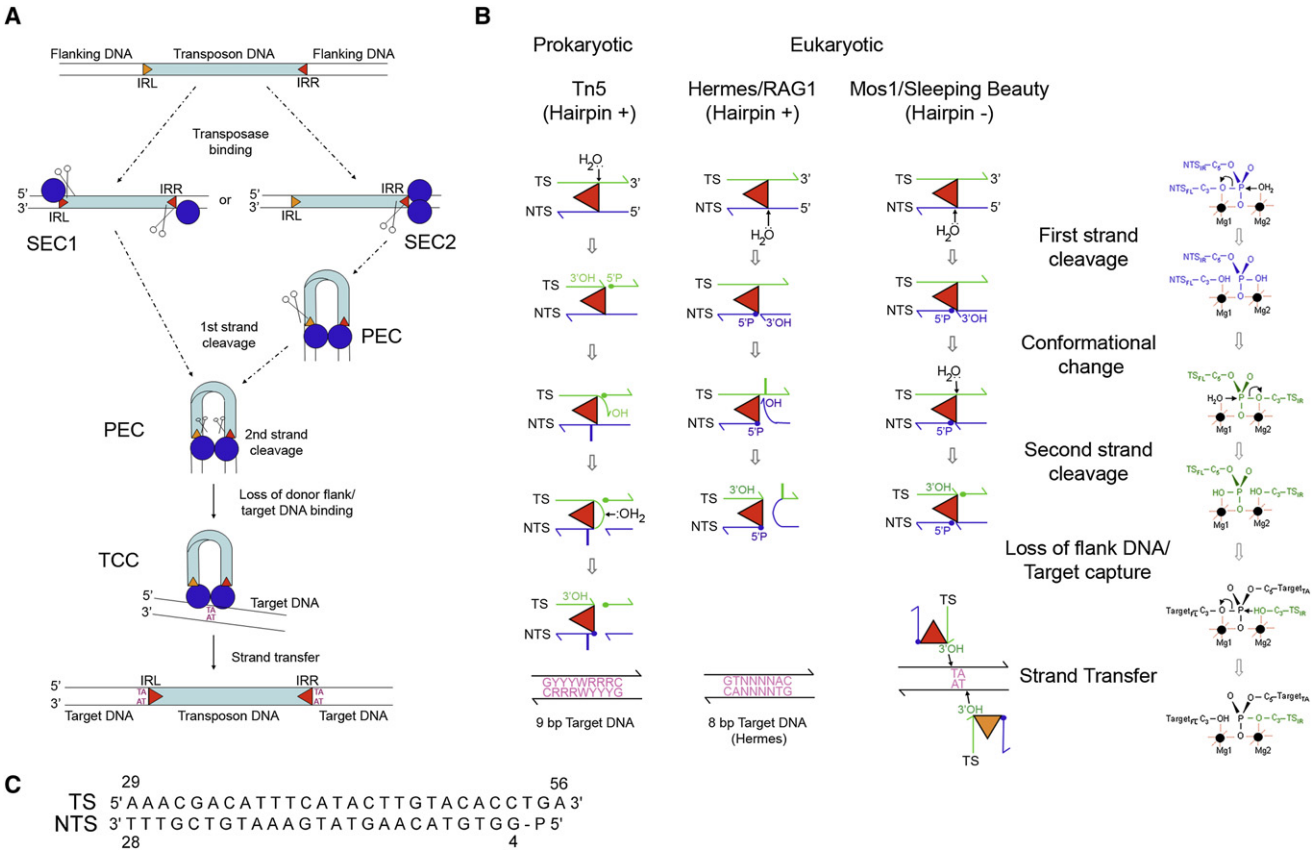


Figure 1. Mechanisms of DNA Transposition

(A) Mos1 transposition pathway: the 1.3kB *Mos1* transposon (light blue) has 28 bp imperfect IRs at both ends (orange triangles) and encodes a transposase (blue circle), the sole requirement for Mos1 transposition. Transposase binds to a single end as a monomer (SEC1) or a dimer (SEC2). The ends are brought together to form a paired-end complex (PEC) and the transposon is excised from flanking DNA. Subsequently target DNA binds, forming the target capture complex (TCC), and the transposon integrates at a TA sequence.

(B) DNA excision proceeds via a hairpin intermediate in transposition of some prokaryotic (Tn5) and eukaryotic (Hermes) elements and flipped bases (shown as thick lines) facilitate hairpin formation (Ason and Reznikoff, 2002; Grundy et al., 2007). By contrast Tc1/*mariner* transposons (e.g., *Mos1* and *Sleeping Beauty*) do not form DNA hairpin intermediates and both strands are presumed to be cleaved by hydrolysis. The 3′OH on the TS attacks target DNA in the strand transfer reaction. Tn5 inserts into a 9 bp consensus sequence, Hermes an 8 bp sequence and Tc1/*mariner* elements always integrate into TA di-nucleotides. In the two metal ion mechanism it is proposed that the roles of the two active site metal ions swap in successive reactions, so that the metal which stabilizes the substrate in one step activates the nucleophile in the next step, and vice versa.

(C) The DNA duplex used for crystallization has the sequence of the right Mos1 IR after DNA excision, with a 3 nucleotide protruding 3′ end.

of excision and integration reactions are different for different elements (Figure 1B). Mu transposase and HIV-1 integrase cleave just one DNA strand at each transposon end, to release the transposon 3′ ends. Other elements excise fully via a DNA hairpin intermediate: the 3′OH released by first-strand cleavage attacks the opposite strand to cleave the DNA and form a hairpin (Figure 1B). This hairpin is on the transposon DNA if transposon 3′ end-cleavage is first (as in Tn5 transposition) or on the flanking DNA if the transposon 5′ end is cleaved first, as in transposition of the eukaryotic hAT elements (Zhou et al., 2004) and V(D)J recombination (Roth et al., 1992). The length of the stagger between the two target DNA sites, into which the 3′ transposon ends are integrated, varies among DDE transposases (Figure 1B). For example, Tn5 transposase inserts into target sites staggered by 9 bp, whereas the separation is 8 bp for Hermes and 5 bp for HIV-1. While Tc1/*mariner* elements have a strict requirement for TA target sequences, other elements show only limited target site sequence preferences, for example Tn5 (Shevchenko et al., 2002) and Hermes (Guimond et al., 2003) (Figure 1B).

Although the catalytic mechanisms of these enzymes have been studied extensively from a biochemical perspective, the structural basis for these differences in reaction specificities remains unknown. The only available structure of an intact DNA transposase in complex with DNA is that of the prokaryotic transposon Tn5 (Davies et al., 2000). Here we report the first structure of a full-length eukaryotic transposase in a paired-end complex with transposon end DNA. It reveals that the ends are held in a *trans* arrangement by a dimeric transposase, in a very different architecture to the Tn5 complex. Two additional DNA duplexes

Table 1. X-Ray Diffraction Data and Refinement Statistics

PEC Crystal	5 mM $MnCl_2$, 5-Iodo-dU	5 mM $MgCl_2$	5 mM $MgCl_2$, 5-Iodo-dU	5 mM $MnCl_2$, 5-Iodo-dU
Unit cell dimensions (Å)	a = 120.8, b = 85.1, c = 132.6, β = 99.3°	a = 121.2, b = 85.0, c = 131.3, β = 98.9°	a = 120.6, b = 85.6, c = 131.8, β = 99.0°	a = 120.8, b = 85.1, c = 132.6,β = 99.3°
Wavelength (Å)	0.934	0.954	0.976	1.894
Resolution (Å)	40.0-3.25 (3.43-3.25)	30.0-3.5 (3.7-3.5)	30.0-4.25 (4.48-4.25)	71.0-4.5 (4.7-4.5)
Total observations	124852 (18489)	128166 (18767)	78964 (11664)	55846 (8182)
Unique observations	38972 (5916)	33149 (4821)	18570 (2719)	15786 (2262)
Rmerge	0.173 (0.445)	0.170 (0.662)	0.187 (0.434)	0.155 (0.567)
Completeness (%)	93.2 (97.6)	98.8 (99.0)	98.5 (99.0)	99.0 (98.7)
$<I/\sigma(I)>$	8.2 (2.5)	9.5 (2.5)	9.4 (3.7)	9.7 (2.8)
Multiplicity	3.2 (3.1)	3.9 (3.9)	4.3 (4.3)	3.5 (3.6)
Anomalous completeness (%)				96.1 (95.5)
Anomalous multiplicity				1.8 (1.8)
Rwork (%)	24.5	21.9		
Rfree (%)	30.1	27.9		
No. molecules per asymmetric unit	1	1		
Vm ($Å^3$ Da^{-1})	5.05	5.10		
Solvent content (%)	75.7	75.9		
rmsd from ideality Bondlength (Å)	0.012	0.015		
Bond angle (deg)	1.70	1.997		
Chirality (Å)	0.101	0.091		
Ramachandran plot core (%)	81.8	73.3		
allowed (%)	15.9	24.7		
generous (%)	2.3	2.0		
disallowed (%)	0.0	0.0		
Average B factor ($Å^2$)	70.0	64.9		
Number of metal ions	1	1		
Number of sulfate ions	2	2		

Data in parentheses are for the highest resolution bin.

in the complex indicate likely binding sites for flanking DNA. Sequence-specific recognition of bases in single-stranded DNA at the IR ends holds the transposon ends in the active sites for target integration. Based on the approximately parallel arrangement of the IRs, and analysis of transposase mutants in strand transfer assays, we present a model for target DNA binding and a structural basis for integration into TA target DNA. First-strand cleavage assays with mixtures of mutant proteins reveal that the transposon 5′ end is cleaved by a transposase dimer. The structure provides insight into other members of the Tc1/*mariner* family of transposons, including elements with promising biotechnology applications such as Sleeping Beauty.

RESULTS AND DISCUSSION

Crystallization of the Mos1 Paired-End Complex

Crystals of the Mos1 paired-end complex (PEC) were formed using full-length transposase, with the mutation T216A that renders the protein soluble, but has no significant effect on catalytic activity (Richardson et al., 2004). This was mixed with double stranded DNA with a 3 nucleotide 3′ overhang mimicking the cleaved Mos1 right transposon end (IRR) (Dawson and Finnegan, 2003), prepared by annealing a 28nt transferred strand (TS) oligonucleotide and a 5′ phosphorylated 25 nt nontransferred strand (NTS) (Figure 1C) (Richardson et al., 2007). Crystals were grown either in the presence of 5 mM $MnCl_2$ or 5 mM $MgCl_2$ and diffracted X-rays to a maximum resolution of 3.25 Å. Crystallographic phases were determined by molecular replacement, using the structure of the Mos1 catalytic domain (2F7T) as the initial search model (see materials and methods). The X-ray diffraction data and refinement statistics are given in Table 1.

The PEC Contains a Dimeric Transposase

The refined crystal structure of the Mos1 PEC is shown in Figures 2A and 2B and represented schematically in Figure 2C. The crystal structure contains a dimer of transposase and four DNA duplexes. Two IRR duplexes (IR DNA_A and IR DNA_B) are recognized by the N-terminal DNA-binding domain of the transposase and are held in position in the catalytic domains as if they have just been cleaved. In striking contrast to the anti-parallel orientation of transposon ends in the Tn5 synaptic complex (Davies et al., 2000), these duplexes are approximately parallel. The N-terminal domain of the transposase (residues 1–112), comprises two helix-turn-helix (HTH) motifs linked by a minor groove

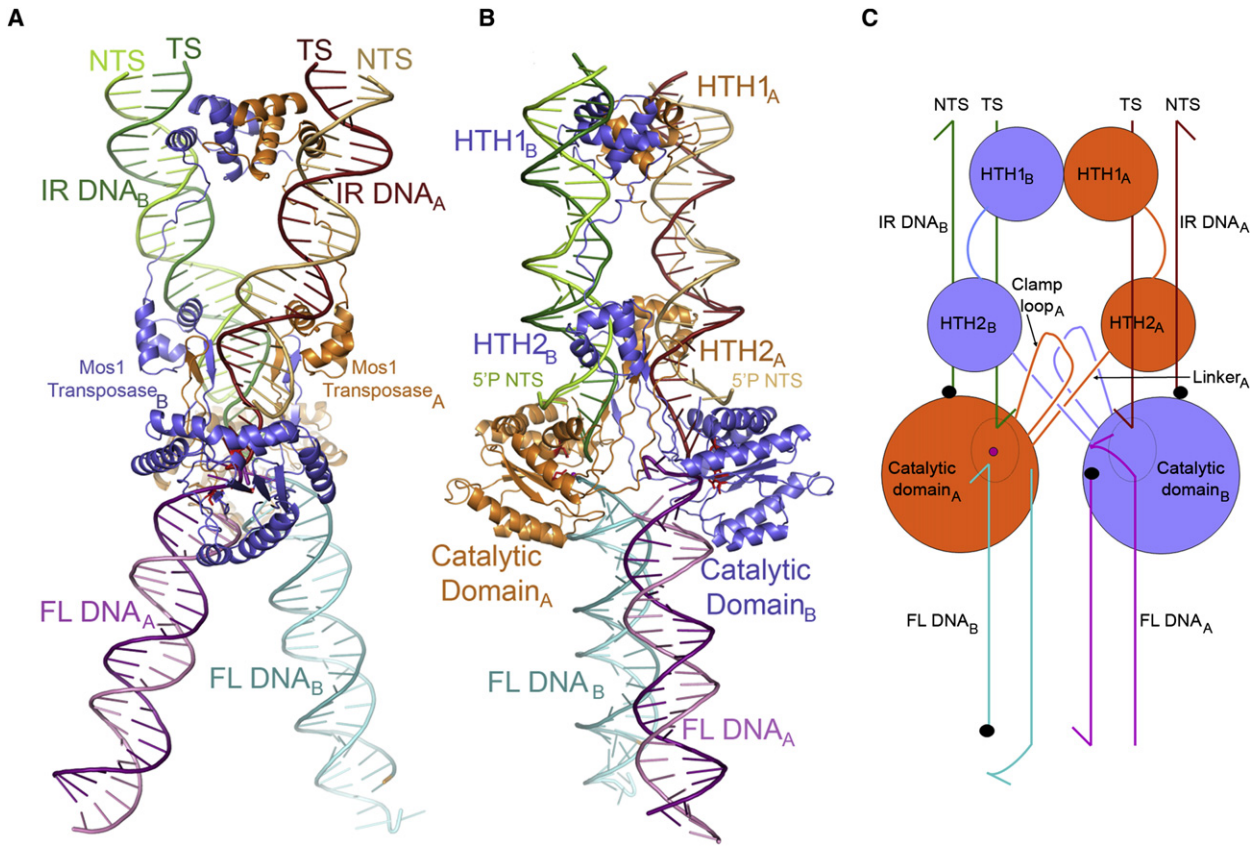


Figure 2. Architecture of the Mos1 PEC

(A and B) Orthogonal views of the PEC crystal structure. Transposase monomer A is colored orange and monomer B blue. The two major-groove DNA-binding motifs contain HTH1 (residues 24–55) and HTH2 (residues 89–110). The minor-groove binding motif comprises residues 63–71. The two DNA duplexes bound by the DNA-binding domains are labeled IR DNA and the two extra DNA duplexes are labeled FL DNA.

(C) Schematic diagram of the structure. An arrow indicates the 3′ end of each DNA strand and a black dot indicates the 5′ phosphate of the NTS. The purple sphere indicates the metal ion in active site A.

binding motif (Figure 2A). Residues 113–125 form a linker between the DNA-binding domain and the catalytic domain (residues 126–161 and 190–345). The catalytic domain has an RNaseH-like fold and is almost identical to our published structure of this domain (Richardson et al., 2006) and can be superimposed with an rmsd of 0.78 Å for all Cα atoms. Residues 162–189 (which were disordered in the catalytic domain structure) form a clamp loop extending out from the catalytic domain making key interactions with the linker of the other transposase monomer in the PEC (Figure 2C). Two additional IRR duplexes (FL DNA$_A$ and FL DNA$_B$) are bound by the catalytic domains in positions that could represent binding sites for DNA flanking the transposon (Figure 2A).

The IR DNA Sequence Is Recognized by the N-Terminal Paired DNA-Binding Domain

The N-terminal 112 residues of Mos1 transposase contain two α-helical motifs (residues 8–53 and 74–110) each of which contains three α helices (Figure 3A); the second and third α-helix of each motif form a helix-turn-helix (HTH) that interacts with the IR DNA major groove. HTH1 binds in the major groove between bases T21 and T26; HTH2 binds in the major groove between bases T8 and G13. There is base-specific recognition of G22 by Arg 48 in HTH1, while Gln 100 in HTH2 interacts specifically with A9. The extended linker between the two helical domains is well ordered. Residues His 65, Gly 66, and Pro 68 bind deep in the AT rich IR DNA minor groove between bases A15 and A18, primarily by shape complementarity. The entire domain architecture has close structural similarity to the bipartite DNA-binding domains of the Tc3 transposase (Watkins et al., 2004) and the PAX6 transcription factor (Xu et al., 1999), and can be superposed on these structures over all Cα atoms with a rmsd of 3.1 Å and 3.5 Å, respectively. The protein-DNA interactions of the N-terminal DNA-binding domain (defined as *cis* interactions) are summarized in Figure S1A, which is available with this article online.

The Transposase Dimer Is Held Together in the PEC by Two Separate Intersubunit Interfaces

The transposase monomers in the PEC make protein-protein contacts in two distinct regions: the clamp loop of one monomer interacts with the linker region of the other monomer (Figures 2C

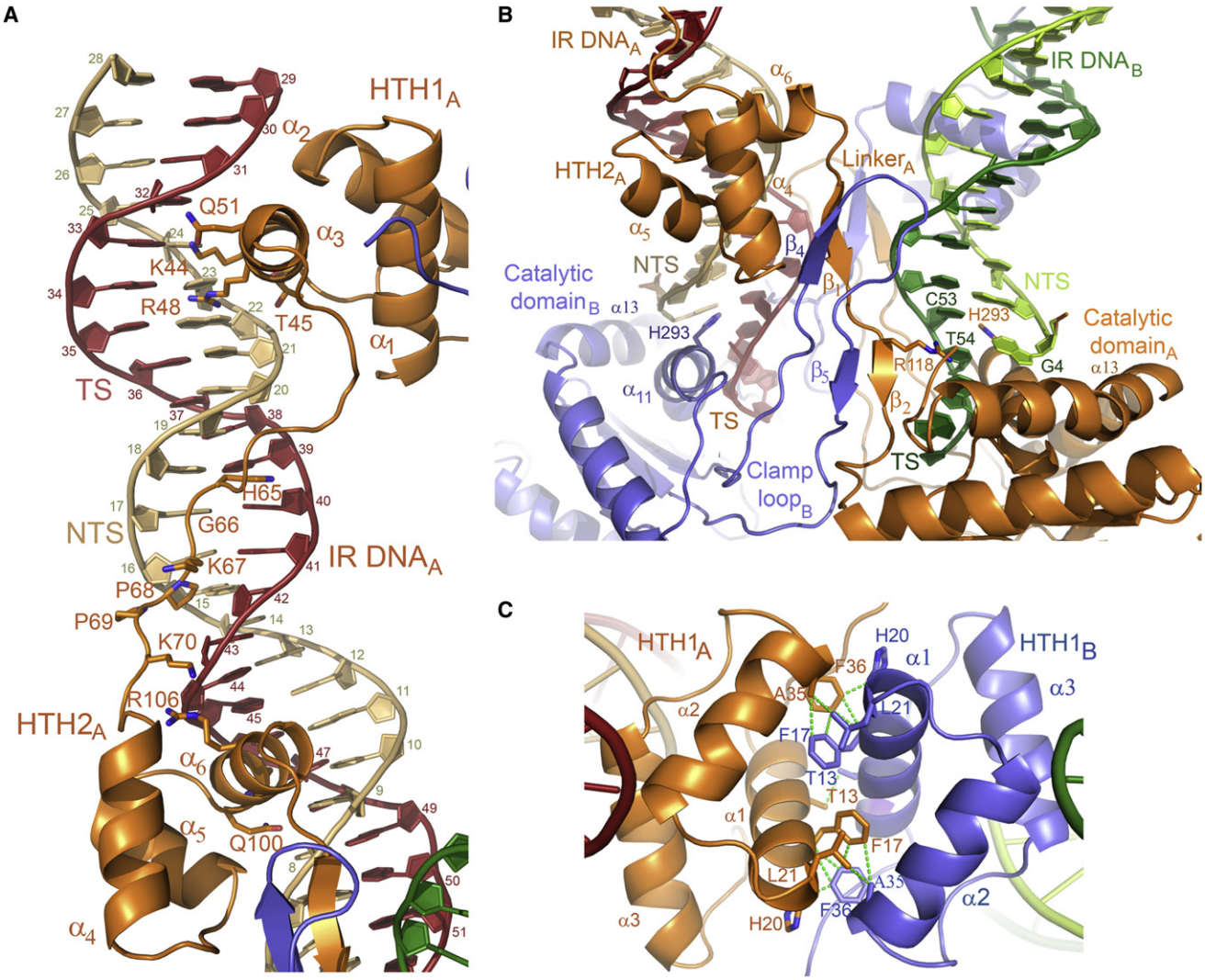


Figure 3. Protein-DNA and Protein-Protein Interfaces

(A) *cis* Protein-DNA interactions between transposase monomer A and IR DNA_A. Transposase is colored orange and shown in ribbon representation. The TS and NTS of IR DNA_A are numbered and colored red and beige respectively. The side-chains of key residues involved in DNA interactions are labeled and shown as sticks.

(B) Interactions between the linker of monomer A (orange) and the clamp loop of monomer B (blue), with the short β strands (β1-4) labeled. There are also symmetry related interactions between the linker of monomer B and the clamp loop of monomer A (data not shown). The staggered ends of IR DNA are lodged on the α11 helices of the catalytic domains and key residues involved in *trans* protein-DNA interactions are labeled.

(C) Dimerisation of the two HTH1 motifs is mediated by hydrophobic contacts (green dotted lines) between α helices 1 and 2 of each transposase monomer. Residues involved in this interface are labeled and shown as sticks.

and 3B), and the HTH1 motifs interact with each other (Figure 3C). These protein dimerization interfaces bury 2146 $Å^2$ surface area in total, consistent with a biologically relevant interaction.

The clamp loop-linker interactions form the major part of the dimer interface, burying 781 $Å^2$ per monomer (1562 $Å^2$ in total). The straight linker between the DNA-binding and catalytic domains (Figure 3B) has two short β strands: β1 (residues 113–116) and β2 (residues 118–120). The clamp loop extends from the catalytic domain, between residues 162–189, and also contains two short β strands: β4 (residues 169–172) and β5 (residues 180–182). A type II β-turn changes the direction of the clamp loop by 180° at Pro 174 and Gly 175. Consequently, β4 of the clamp loop of one transposase monomer forms an antiparallel β sheet with β1 of the linker of the other monomer, while β5 and β2 form a parallel β sheet.

The interface between the two HTH1 motifs holds the inner ends of the IR DNA duplexes together in the PEC (Figures 2A, 2B, and 3C). Residues in α helices 1 and 2 (Thr13, Phe17, His20, Leu21, Ala35, and Phe36) form extensive hydrophobic contacts across the two-fold noncrystallographic symmetry axis. This interface buries 584 $Å^2$ surface area, and is similar in architecture to the dimer interface observed in the crystal

structures of the Tc3 DNA-binding domain in complex with IR DNA (van Pouderoyen et al., 1997; Watkins et al., 2004). This dual role for a small HTH motif, in mediating both sequence-specific DNA binding and synapsis of two DNA sites, was also observed in the synaptic complex of Sin recombinase (Mouw et al., 2008).

The dimer interface of the catalytic domains in the PEC is different from that observed in the crystal structure of the catalytic domain in the absence of DNA (Richardson et al., 2006) (compare Figures S2A and S2B); some of the residues implicated in the latter crystallographic interface (H293, W119-E123, K190, and E345) are close to protein-DNA interfaces in the PEC, suggesting that competition between transposase DNA recognition and transposase dimerization (in the absence of DNA) may contribute to regulation of transposition by a transposase concentration dependent mechanism (Lohe and Hartl, 1996).

The Transposase Dimer Locks the Transposon Ends in a *trans* Arrangement

There is a requirement for pairing of the ends in a synaptic complex for TS cleavage in *Mos1* excision (Dawson and Finnegan, 2003). This may prevent undesirable chromosome breaks at a single end, before the transposon is committed to the transposition pathway. The architecture of the *Mos1* PEC provides a molecular explanation for this requirement. The complex is in a *trans* arrangement: that is, the IR DNA bound sequence-specifically by the DNA-binding domain of monomer A interacts with the catalytic domain of transposase monomer B, and vice versa (Figures 2B and 2C), thus both monomers act together to carry out cleavage. The protein cross-over is provided by the linker (residues 113–125) joining the DNA-binding and catalytic domains, and is clearly visible in the electron density map.

The interactions of the catalytic domain with the IR DNA are focused on the outer end of the IR DNA close to the cleavage site (summarized in Figure S1B). The minor groove at the staggered IR DNA end is lodged onto α-helix 11 (residues 287–297), and His 293 disrupts the base-pair hydrogen bonds between C53 and G4, the final base-pair in the IR DNA duplex (Figure 3B). There are further *trans* interactions between residues in the linker and the TS of IR DNA.

trans catalysis has also been demonstrated for Mu (Savilahti and Mizuuchi, 1996) and Tn5 (Naumann and Reznikoff, 2000) transposases and RAG1 recombinase (Swanson, 2001), and a *trans* arrangement was observed in the structure of the Tn5 synaptic complex (Davies et al., 2000). The *trans* arrangement in the Mos1 PEC suggests that this is a recurrent feature in DNA transposition and related genome rearrangements, despite the completely different architectures of the Mos1 and Tn5 transpososomes.

Additional DNA Duplexes Indicate Putative Binding Sites for Flanking DNA

The Mos1 PEC in the crystals was formed with an IRR duplex corresponding to precleaved transposon DNA, rather than by cleavage of a precursor substrate. Surprisingly, there are two additional IRR duplexes (FL DNA_A and FL DNA_B) interacting with each PEC in the crystal. The FL DNA duplexes have the same sequence as IR DNA but are not bound by the transposase N-terminal DNA-binding domains. Instead they interact nonsequence-specifically with the catalytic domains of the transposase (Figure S3). Each FL DNA duplex links two separate PECs together in the crystal lattice, and is presumed to have been captured from the excess DNA in solution during crystallization (Richardson et al., 2007), thereby enabling crystal packing.

The FL DNA duplexes are related by a crystallographic two-fold screw symmetry operation (Figure S4A); the crystallographic asymmetric unit contains only three copies of the DNA duplex (IR DNA_A, IR DNA_B and FL DNA_A) plus the transposase dimer, while FL DNA_B belongs to a neighboring asymmetric unit. The DNA bound to the catalytic domain of monomer A (labeled FL DNA_B) has the blunt end in the active site (Figure 4A), whereas the DNA duplex in the catalytic domain of monomer B (FL DNA_A), has the staggered end in the active site (Figure 4B). This asymmetric arrangement breaks the near perfect two-fold (noncrystallographic) symmetry of the PEC. The single orientation of each FL DNA with respect to its catalytic domain was confirmed by the observation of one signal from iodinated DNA, with T16 replaced by 5-Iodo-dU (Figure S4B), and clear electron density for the 3 overhanging bases of the TS and the terminal 5′ phosphate of the NTS in a Fo-Fc map.

Mos1 transposase makes nonspecific contacts with the backbone phosphates of FL DNA, via residues in the loops formed by residues Thr 213 to Ala 216 and Asn 250 to Arg 257 (summarized in Figure S1C). In addition, Pro 252 is inserted into the FL DNA minor groove. Furthermore, the aromatic ring of Phe 187 stacks against the purine ring of the bases at the 5′ end of each FL DNA strand (A29 and G4) and the 3′OH of T28 on FL DNA_B is coordinated to the divalent metal ion in active site A (Figures 4A and 4C). These nonsequence-specific transposase:DNA interactions and the positions of FL DNA, in-line with IR DNA (Figure 2A), support the proposal that these duplexes are in sites that would be occupied by flanking DNA in the PEC. The position of the FL DNA 3′ OH in active site A, coordinated to Mg1, is consistent with its production by cleavage of a NTS in this active site. The difference in polarity between FL and IR DNA suggests a rearrangement of the catalytic domains with respect to the IR sequences in order to cleave the TS.

Active Site A Contains One Divalent Metal Ion

Active sites A and B, formed by D156, D249, and D284 of monomer A and B, respectively, have similar architectures to the active site in the catalytic domain structure (Richardson et al., 2006) (Figures 4A and 4B). That structure was determined from crystals grown in 5 mM $MgCl_2$ and the active site contained one Mg^{2+} ion coordinated to D156 and D249. We also showed that in the absence of DNA and in the presence of 20 mM $MnCl_2$, a second Mn^{2+} ion was bound in the active site, by D156 and D284 (Richardson et al., 2006). The ability to bind two Mn^{2+} ions is consistent with a two-metal mechanism for catalysis (Beese and Steitz, 1991; Yang et al., 2006) which would allow a transposase dimer to perform all three steps of Mos1 transposition: first-strand cleavage, second strand cleavage and integration into target DNA (Figure 1).

The PEC crystals were grown either in the presence of 5 mM $MgCl_2$ or 5 mM $MnCl_2$ and the DNA-bound structures contain

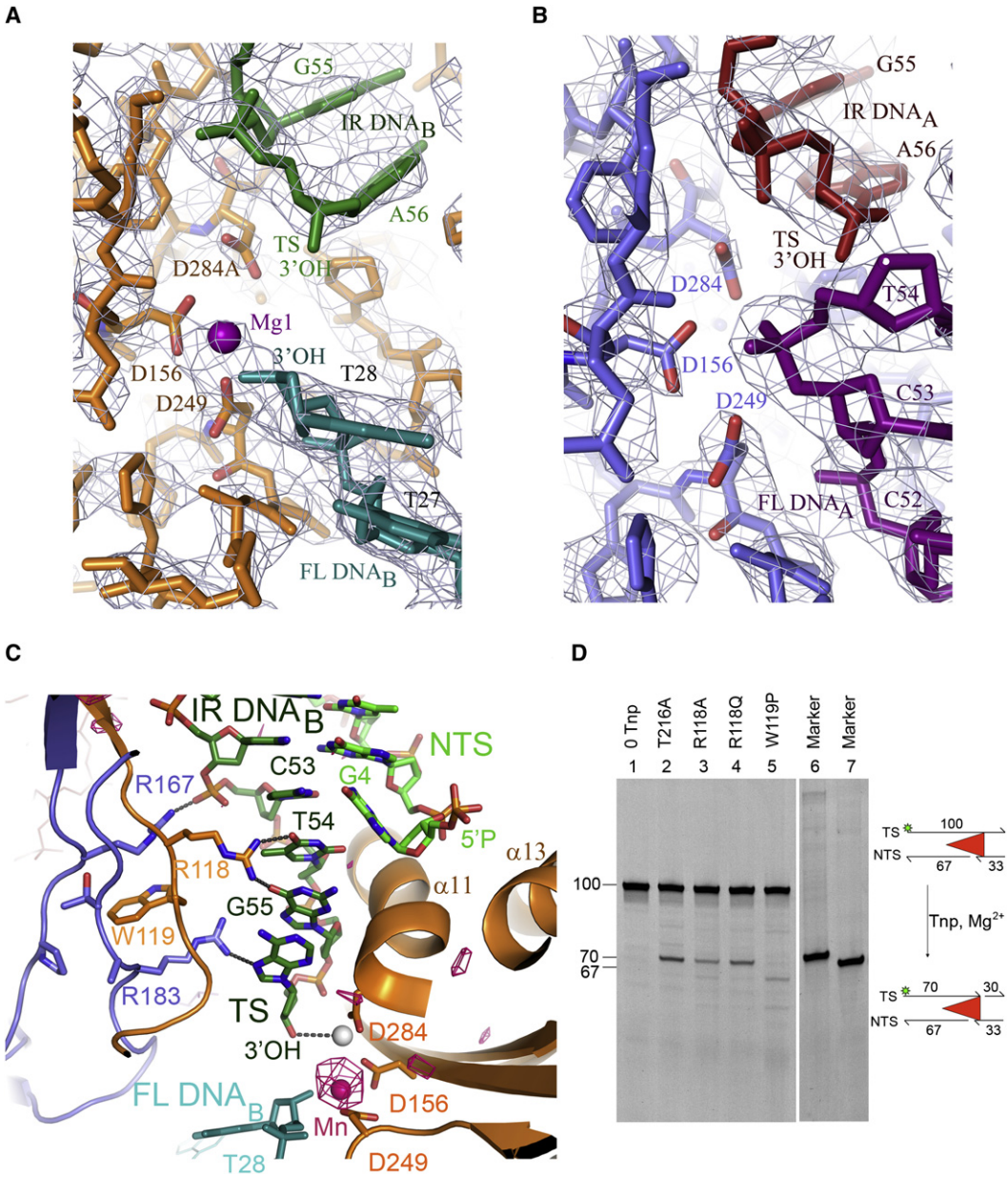


Figure 4. The Active Sites

(A) The active site of monomer A has one metal ion bound in site 1: the position of the Mg^{2+} ion is shown as a purple sphere.

(B) Active site of transposase monomer B does not contain a metal ion. The 2Fo-Fc electron density map (gray mesh) is contoured at 1.7σ.

(C) Key interactions between residues in the linker, the clamp loop and the single-stranded bases at the 3′ end of the TS. A Mn anomalous difference density map (pink mesh, contoured at 3.5σ) confirms the position of the single Mn^{2+} ion in active site A in the Mn^{2+} bound PEC structure (pink sphere). The position of the second (unoccupied) metal binding site (Richardson et al., 2006) is shown as a gray sphere.

(D) Mutation of residues R118 and W119 in the linker region diminished transposase activity in second strand cleavage assays. Cleavage of the fluorescein labeled 100 nt TS was detected by observation of the 70 nt reaction product on an 8% denaturing polyacrylamide gel. Lanes 6 and 7 contained fluorescein labeled DNA markers of 70 and 67nts respectively.

one Mg^{2+} or Mn^{2+} respectively, in active site A, coordinated to D156, D249, and the terminal 3′OH of T28 of FL DNA_B (Figure 4A). However, there is no metal ion in active site B (Figure 4B). A Mn anomalous difference electron density map (shown in Figure 4C as a pink mesh) confirmed the presence of one Mn^{2+} in active site A and the absence of Mn^{2+} in the other potential metal-ion binding sites. The presence of only one metal ion in active site A may be due to the low divalent metal ion concentration in the crystallization conditions. The differences in metal binding in the two active sites can be accounted for

by the presence of the staggered-end of FL DNA_A and the phosphate backbone atoms of T54 in active site B, which appear to preclude metal ion binding in this active site (Figure 4B).

Sequence-Specific Recognition of Unpaired Bases Positions the Transposon 3′ End in the Active Site

Cleavage of the TS generates the transposon 3′OH that is the nucleophile for the subsequent target integration reaction. This must be precisely at the end of the IR to prevent shortening or elongation of the transposon at each transposition event. In the PEC structure, the terminal 3′OH of the TS of IR DNA is bound in the active site, less than 4 Å from the unoccupied metal-binding site (Richardson et al., 2006) formed by D284 and D156 (Figures 4A and 4B), and is ideally placed to act as a nucleophile in the subsequent strand transfer step.

The staggered end of each IR DNA is held in position in the catalytic domain by extensive contacts with both transposase monomers (Figures S1A and S1B). The 5′ end of the NTS, which is three bases inside the IR, is remote from the active site, on the top face of the catalytic domain, close to α-helix 13 (residues 319–328) (Figure 4C) suggesting displacement of this strand out of the active site after first-strand cleavage. The specific sequence of the three unpaired bases on the TS, generated by the staggered cleavage, is read by key transposase-DNA interactions in the PEC: Arg 118 (on the linker between DNA-binding and catalytic domains) forms two hydrogen bonds in *trans* with T54 and G55 (Figure 4C). Mutation of these bases from TG to CC (where no such hydrogen-bonds could form) strongly inhibited second strand cleavage and PEC formation, but did not significantly affect first-strand cleavage (Dawson and Finnegan, 2003). Consistent with the role of Arg 118 in positioning the TS, the double mutant transposase R118Q/T216A diminished the initial rate of second strand cleavage (to ~60% of the activity), whereas R118A/T216A reduced the initial rate of second strand cleavage even further to ~30% of T216A (Figure 4D). The T216A mutation in these double mutants enables soluble expression of transposase in *E. coli* and has no significant effect on transposase activity. The terminal base of the IR DNA (A56) is recognized by a *cis* interaction between an NH_2 group of R183 (on the clamp loop) and N7 of the purine ring; this interaction could also form with G56, the terminal nucleotide of the TS on the left end of Mos1. Furthermore Trp 119 in the linker forms protein-protein interactions with residues of the clamp loop (R167 and R183); the importance of these interactions and flexibility in this region is demonstrated by the phenotype of the mutation W119P, which abolished normal second strand cleavage (Figure 4D) and produced ectopic cleavage products, smaller than the expected 70-mer, indicating that this mutation leads to aberrant positioning of the IR DNA in the active site. Together the transposase-IR DNA interactions position the TS so that the terminal 3′OH is engaged in the active site (Figures 4A and 4B) where it is poised to carry out the next step of transposition: strand transfer into target DNA.

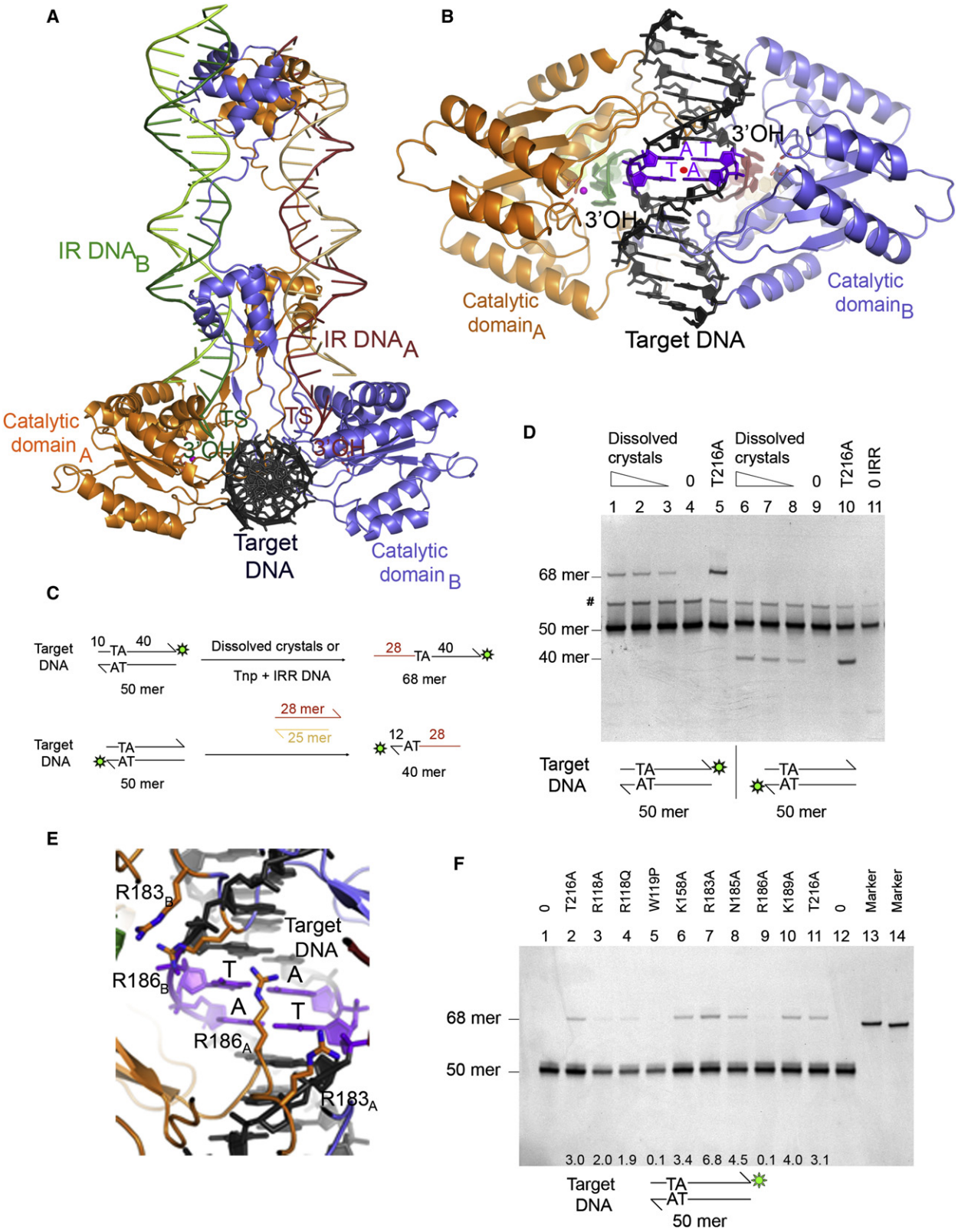
A Model for the Target Capture Complex

Modeling the removal of the FL DNA from the PEC reveals a channel between the catalytic domains into which B-form target DNA can be docked so that the TA target sequence is in close proximity to both active sites (Figures 5A and 5B). The orientation of the modeled target DNA is approximately perpendicular to both IR DNA molecules, and it is bound to the two transposase catalytic domains so that the TA target sequence is positioned on the axis of two-fold symmetry in the PEC (Figure 5B). The 3′ hydroxyls of the transferred strands are separated by 24.6 Å across the dimer interface of the Mos1 PEC, and each is 5.8 Å away from the respective scissile phosphate on the modeled target DNA. The parallel orientation of the two IR DNA molecules is ideal for integration of the transferred strands into the opposite strands of B-form target DNA with 2 bp spacing between the insertion sites (corresponding to a distance of 18 Å between the two scissile phosphates). Interestingly, the very different architecture of the Tn5 synaptic complex, which has an anti-parallel arrangement of the transposon ends, places the TS 3′ hydroxyls in the correct orientation for in-line attack of target DNA at sites separated by 9 bp (Davies et al., 2000), assuming that target DNA is in the B-form. The arrangement of transposon ends in the paired-end complexes of other elements will presumably also reflect the characteristic staggered positions of insertion of that element.

In a strand transfer assay (Figure 5C) dissolved crystals integrated the IR DNA that they contained into fluorescently labeled target DNA substrates (Figure 5D) demonstrating that the Mos1 PEC in the crystals can carry out strand transfer. The proximity of the clamp loop to the TA di-nucleotide of the modeled target DNA suggested five residues which could potentially be involved in target DNA recognition: K158, R183, N185, R186 and K189, where R186 is the residue closest to the TA sequence (Figure 5E). To test the target DNA binding model (Figure 5), we mutated each of these residues, individually, to alanine and measured the activity of the mutant transposases in DNA cleavage and strand transfer assays; all mutant transposases also contained the T216A solubilising mutation. R186A was fully competent in second strand cleavage assays (Figure S5) but showed a dramatic reduction in strand transfer, to less than 5% of the T216A transposase (Figure 5F). By contrast the mutants K158A, N185A, and K189A showed a slight increase in target integration, whereas R183A showed a more significant increase; none of these mutants had any detectable effect on second strand cleavage. We propose that together the two R186 side-chains play a key role in binding the symmetrical TA target DNA sequence (Figure 5E), possibly by recognition of the TA base pairs or the TpA dinucleotides (Luscombe et al., 2001).

The Mos1 Transposition Pathway

The *trans* arrangement of the transposase catalytic and DNA-binding domain in the Mos1 PEC structure, with the TS 3′ end in the active site, strongly indicates that second strand cleavage occurs in *trans* within a dimeric PEC. Without a hairpin intermediate during DNA excision, there must be a conformational change in the transpososome between the first and second strand cleavages, to remove the NTS from the active site and position the TS correctly for cleavage (Figure 1B). The position of the 5′ end of the NTS in the PEC, remote from the active site, is consistent with displacement of this strand after first-strand cleavage.


A
IR DNA B
IR DNA A
Catalytic domain A
TS
TS
3'OH
3'OH
Target DNA
Catalytic domain B
B
3'OH
3'OH
Catalytic domain A
Target DNA
Catalytic domain B
C
Target DNA
10
TA
40
AT
50 mer
Dissolved crystals or
Tnp + IRR DNA
28
TA
40
68 mer
28 mer
25 mer
Target DNA
TA
AT
50 mer
12
AT
28
40 mer
D
Dissolved crystals
0
T216A
Dissolved crystals
0
T216A
0 IRR
1 2 3 4 5 6 7 8 9 10 11
68 mer
#
50 mer
40 mer
Target DNA
TA
AT
50 mer
TA
AT
50 mer
E
R183B
Target DNA
R186B
T
A
A
T
R186A
R183A
F
0
T216A
R118A
R118Q
W119P
K158A
R183A
N185A
R186A
K189A
T216A
0
Marker
Marker
1 2 3 4 5 6 7 8 9 10 11 12 13 14
68 mer
50 mer
3.0 2.0 1.9 0.1 3.4 6.8 4.5 0.1 4.0 3.1
Target DNA
TA
AT
50 mer

First-strand cleavage can occur prior to pairing of the ends (Dawson and Finnegan, 2003; Lipkow et al., 2004), within a single-end complex (SEC) comprising one IR and either a monomer (SEC1) or dimer (SEC2) of transposase (Auge-Gouillou et al., 2005) (Figure 1A).

NTS cleavage in SEC1 would be necessarily in *cis* and cleavage at both ends could occur prior to synapsis, with the PEC formed by dimerization of SEC1 (Figure 1A). Alternatively, NTS cleavage in SEC2 could be *cis* or *trans* (Figure 6A) with sequential cleavage of the two ends: nicking of one end within SEC2 and the other within a PEC.

To distinguish these possibilities we performed NTS cleavage assays with mixtures of Mos1 transposase mutants (Figure 6). If NTS cleavage is in SEC2, heterodimers formed in a mixture of a DNA-binding deficient mutant and a catalytically inactive mutant would be active if cleavage is in *trans* but inactive if cleavage is in *cis* (Figure 6A); homodimers in the mixture would be inactive. Conversely, heterodimers formed in a mixture of T216A transposase and a mutant deficient in both DNA binding and catalysis would be active if NTS cleavage is in *cis*, but inactive if cleavage is in *trans*. However, T216A homodimers in this mixture are also active. If NTS cleavage is by a monomer in SEC1, mixing T216A transposase with an inactive transposase would have no effect on T216A cleavage activity.

We prepared a DNA binding deficient Mos1 transposase by mutating R48 in HTH1 to Gln and Q100 in HTH2 to Arg. A catalytically inactive transposase was created by mutating the third residue of the DDD motif (D284) to Ala. A DNA binding and catalytically inactive transposase incorporating all three mutations (R48Q, Q100R and D284A) was also created. Each mutant transposase also contained the solubilising mutation T216A and was purified as previously described (Richardson et al., 2004) (Figure S6A). Gel filtration chromatography confirmed that each mutant was a dimer at a concentration of 1 μM (data not shown). The R48Q/Q100R and D284A mutants had the expected properties when analyzed for DNA-binding (Figure S6B) and NTS cleavage (Figure 6B).

First we mixed the R48Q/Q100R and D284A mutants in varying ratios, with the total protein concentration constant at 30 nM. In NTS cleavage assays, no cleavage product was observed (Figure S7A). This result could be interpreted as a *cis* arrangement of transposase and IR DNA in SEC1 or SEC2, lack of heterodimer formation or formation of incompetent heterodimers. To confirm heterodimer formation, we then mixed each mutant with T216A transposase over a range of ratios (with total protein concentration constant at 30 nM). The activity of the mixtures was compared with that of T216A transposase in dilution buffer at concentrations equivalent to those in the mixtures. Mixing the DNA-binding deficient mutants R48Q/Q100R or R48Q/Q100R/D284A with T216A transposase reduced NTS cleavage activity compared to the control (Figure 6C, S7B, and S7C). Similarly, titration of these mutants into fixed concentrations of T216A transposase (30 nM or 5 nM) reduced NTS cleavage activity (Figure 6D). Because these mutants cannot bind DNA, the observed inhibition of cleavage must be due to mixed multimers forming with T216A and is consistent with NTS cleavage by a transposase dimer.

Heterodimers formed between T216A and R48Q/Q100R transposases are predicted to be active whether arranged in *cis* or *trans* (Figure 6A). However, at an equimolar ratio of T216A and R48Q/Q100R transposases (15 nM of each) NTS cleavage was reduced to ~29% of the activity of T216A transposase at 30 nM (Figure 6C). As most of this activity can be attributed to T216A homodimers (predicted to represent 25% of the mixture), we conclude that T216A:R48Q/Q100R heterodimers have negligible activity. This could also explain the lack of activity of D284A:R48Q/Q100R heterodimers (Figure 6C), and so no conclusions can be drawn regarding the *cis* or *trans* arrangement of the transposase in SEC2 from these results.

The catalytically inactive mutant D284A was mixed with T216A transposase over a range of ratios (with total protein concentration constant at 30 nM). At an equimolar ratio of transposases, NTS cleavage was reduced to 46% of the activity of T216A transposase at 30 nM (Figures 6C and S7B). This is consistent with formation of T216A:D284A heterodimers with half the activity of T216A homodimers and the dimer model for cleavage. Titration of the D284A mutant into a fixed concentration of T216A transposase (30 nM) reduced NTS cleavage due to competition for the DNA substrate by the inactive transposase. (Figure 6D).

CONCLUSIONS

Transposition of other Tc1/*mariner* elements is mechanistically similar to *Mos*1: DNA excision generates a staggered double-strand break at the transposon ends, without formation of a DNA hairpin intermediate, and most elements insert into a TA target sequence (Plasterk et al., 1999). Comparison of the structures of the Mos1 PEC and the Tc3 DNA-binding domain:DNA complex (Watkins et al., 2004) with the domain arrangements predicted from the sequences of Sleeping Beauty and SETMAR

Figure 5. Target DNA Binding Model

(A) Proposed target DNA (black) binding site. The TS terminal 3′OHs are marked.

(B) View from the underside of the PEC, with the target TA sequence (purple) and the TS 3′OHs highlighted. The 2-fold axis of symmetry is marked (red dot).

(C) Strand transfer assay. The 50-mer target DNA substrates contained one TA dinucleotide, 10 and 11 bases from one end, and a 3′ fluorescein tag on either the top or the bottom strand. Integration of the 28-mer TS of IRR DNA yields a 68-mer or 40-mer labeled product depending on whether the top or bottom strand of target DNA was labeled.

(D) Dissolved crystals integrate into target DNA (lanes 1–3 and 6–8) to give identical integration products to T216A transposase plus IRR DNA in solution (lanes 5 and 10). Negative controls: without crystals (lanes 4 and 9) or with T216A but without IRR DNA (lane 11). The band labeled # corresponds to an alternative secondary structure of the target DNA substrate which was not fully denatured on this gel.

(E) View of the target TA sequence (purple) from the topside of the PEC, with the positions of R186 and R183 in each monomer highlighted.

(F) Target integration assay of double mutant transposases (lanes 3–10) compared with the activity of T216A transposase (lanes 2 and 11). Lanes 1 and 12 correspond to reaction without transposase and lanes 13 and 14 are 67 nt and 66 nt DNA markers, respectively. Percentage strand transfer is indicated at the bottom of each lane.

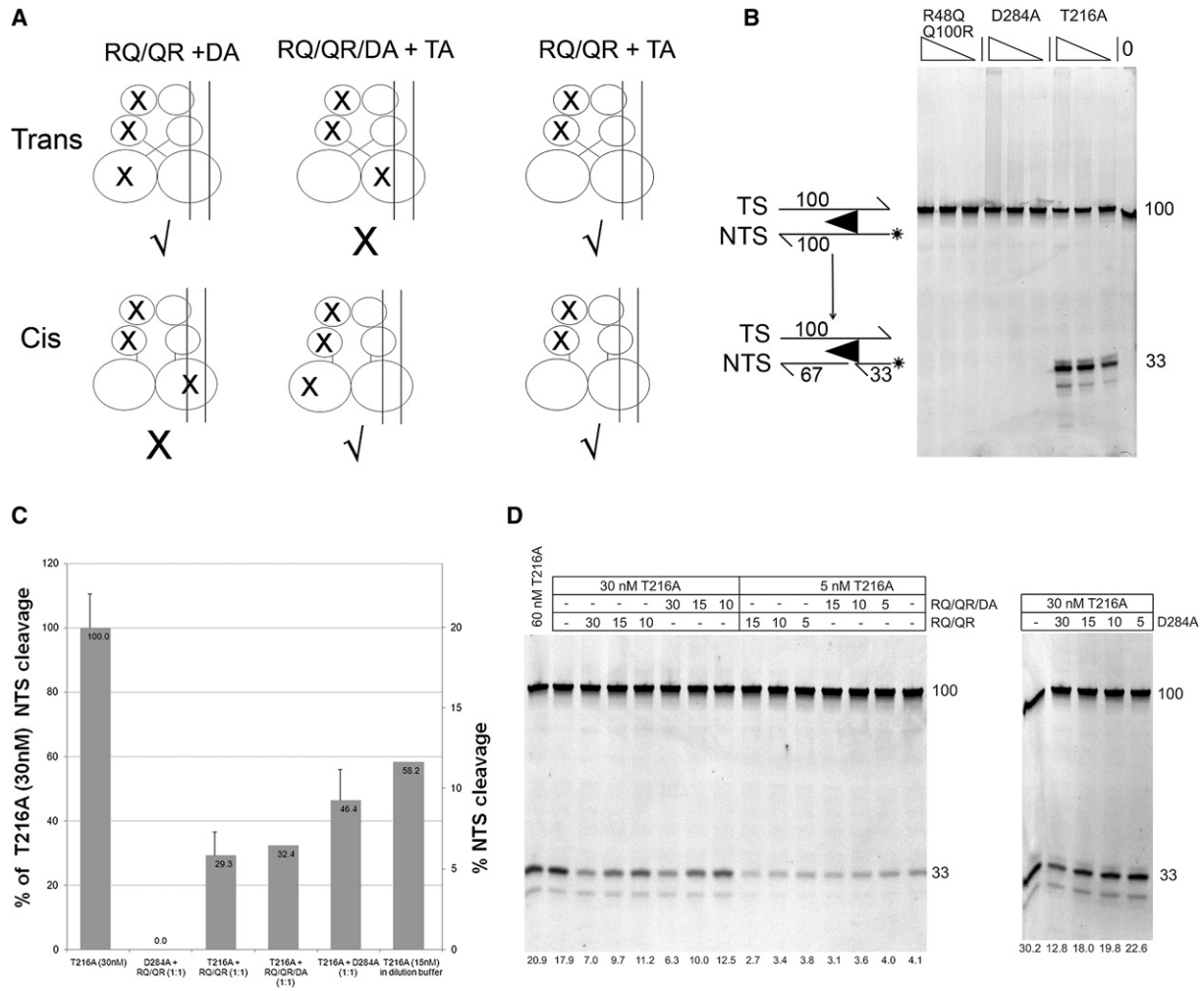


Figure 6. First-Strand Cleavage Assays

(A) Schematic of the possible *trans* or *cis* arrangements of transposase heterodimers in SEC2 complexes and their predicted cleavage activity.

(B) Cleavage assay of T216A transposase and R48Q/Q100R and D284A mutants at 60 nM, 30 nM, and 15 nM.

(C) Graph of percentage cleavage of 1:1 mixtures of mutant transposases in assays where the total protein concentration was constant at 30 nM. Data are represented as mean ± SD.

(D) Mixed mutant cleavage assay in which mutants were titrated into fixed concentrations of T216A transposase (30 nM or 5 nM). The percentage cleavage activity is shown at the bottom of each lane.

(Cordaux et al., 2006) highlights the similarities in Tc1/*mariner* transposases (Figure S8). There are only small variations in length of the clamp loop and the sequence of this motif displays some conservation in the Tc1/*mariner* family. It seems likely that this loop is a conserved structural feature that plays a role in dimer PEC formation and target DNA binding throughout the Tc1/*mariner* family.

The general mechanism outlined here, based on the Mos1 PEC structure, provides a template for the whole family. The *trans*, dimeric architecture of the Mos1 PEC provides molecular insight into many facets of the Mos1 transposition mechanism: sequence-specific recognition of the transposon ends, the requirement for pairing of the ends prior to TS cleavage, putative binding sites for flanking DNA and key molecular interactions that position the TS 3′OH for subsequent target DNA integration. A model for the target capture complex, supported by biochemical analysis of mutant transposases, reveals how the parallel arrangement of IR DNA in the Mos1 PEC facilitates integration of the transposon ends into positions separated by only 2 bp in target DNA. Biochemical experiments with mixtures of mutant proteins reveal that the NTS is cleaved by a dimer of transposase. The structure will guide further biochemical experiments on Mos1 transposition and increase our understanding of DNA transposition in general.

EXPERIMENTAL PROCEDURES

Structure Determination and Refinement

Crystals were grown as described previously (Richardson et al., 2007). All crystals displayed monoclinic ($P2_1$) symmetry. Data sets were collected from three

Mos1 PEC crystals: native crystals grown in 5 mM $MgCl_2$ and two iodinated DNA derivative crystals, with T16 replaced by 5-Iodo-dU, grown in 5 mM $MnCl_2$ or 5 mM $MgCl_2$ (Table 1). An anomalous dataset of the $MnCl_2$ containing PEC crystal was collected at a wavelength of 1.89 Å to confirm the position of the Mn^{2+} ion.

Initial phases were determined by molecular replacement using our structure of the catalytic domain (2F7T) as the search model in PHASER followed by further molecular replacement with a 20 bp B-DNA duplex. The remaining structure was built manually. Restrained refinement was performed with Refmac and included weak noncrystallographic symmetry restraints on the protein atoms and TLS restraints. The electron density is well defined apart from the N-terminal four residues of the transposase and residues 239 to 242 in the catalytic domain. Structure diagrams were prepared using PyMOL (http://pymol.sourceforge.net/).

Preparation of Mutant Transposases

Each mutant transposase was cloned with the QuikChange site-directed mutagenesis kit (Stratagene), using the T216A plasmid and a pair of primers designed to contain the required additional mutation (synthesized by Integrated DNA Technologies (IDT), Iowa), according to the manufacturer's protocol. Mutagenesis was confirmed by DNA sequencing of resulting plasmids.

Second-Strand Cleavage Assay

The dsDNA substrate, precleaved at the site of first-strand cleavage, was prepared by annealing a fluorescently labeled 100 nt TS with 67-mer and 33-mer NTS oligonucleotides. All oligonucleotides were purified by ion exchange HPLC or PAGE by the manufacturer (IDT) and the labeled 100 nt TS was re-purified by PAGE. The 100-mer TS (5′ TTT CTT TTT CCA CAA AAT TTA ACG TGT TTT TTG ATT TAA AAA AAA CGA CAT TTC ATA CTT GTA CAC CTG Atagtttctatattcaccgactggagcccgt) contained the terminal 70 nt of Mos1 transposon DNA (including the 28 nt right-hand IR sequence (underlined), 30 nt of flanking DNA (lower case) and a 6-carboxyfluorescein tag at the 5′ end). The 67 nt and 33 nt NTS had the sequences 5′ GGT GTA CAA GTA TGA AAT GTC GTT TTT TTT AAA TCA AAA AAC ACG TTA AAT TTT GTG GAA AAA GAA A and 5′-acg ggc tcc agt cgg tga ata tag aaa cta TCA respectively.

Oligonucleotides were annealed by heating a 10% excess of the 67 nt and 33 nt unlabelled NTS with fluorescently labeled TS and heating to 85°C for 5 min in 10 mM Tris buffer, 50 mM NaCl and 1 mM EDTA before cooling to 25°C over a period of 1 hr.

For the cleavage assay, 30 nM dsDNA substrate was incubated with 25 nM transposase in a final volume of 20 μl for 40 min at 30°C in buffer containing 25 mM HEPES/NaOH pH7.5, 50 mM Potassium Acetate, 10% (*v/v*) glycerol, 0.25 mM EDTA, 1 mM DTT, 10 mM $MgCl_2$, 50 μg/mL BSA and 20% (*v/v*) DMSO. The reaction was stopped by the addition of 20 μl of loading dye (95% (*v/v*) formamide, 20 mM EDTA, 0.05% (*w/v*) bromophenol blue, 0.05% (*w/v*) xylene cyanol) and the products separated on an 8% denaturing polyacrylamide gel containing 7.5 M urea and 1× TTE running buffer (89 mM tris base, 29 mM taurine and 0.5 mM EDTA) at ~20 V/cm to heat the gel to approximately 50°C. To visualize the products, 6-carboxyfluorescein was excited using the 473 nm laser and detected with the 510 nm long pass filter on a Fuji BAS-5000 system. Fluorescence was quantified using Image Gauge 4.0 software using peak area methods. Time course experiments, with reactions stopped at 5, 10, 20, and 60 min, were used to estimate the initial reaction rates.

Target Integration Assays

The 50-mer target DNA substrates contained one TpA dinucleotide and were prepared by annealing the 50 nt top strand (5′AGC AGT GCA CTA GTG CAC GAC CGT TCA AAG CTT CGG AAC GGG ACA CTG TT) with the complementary bottom strand; the 3′ end of either the top strand or the bottom strand was labeled with a fluorescein tag. Oligonucleotides were purified by IE HPLC by the manufacturer (IDT), and purified by PAGE and annealed as above. The IR DNA substrate (IRR) had the same composition as the dsDNA used for the crystallization except that the NTS was not 5′ phosphorylated (28 nt TS annealed to the 25 nt NTS).

Crystals were washed three times in well solution before being dissolved in 20 μl of distilled water. The final concentration of protein in the dissolved crystals was estimated to be 6 μM.

Target integration assays of transposase mutants were performed in 20 μl reactions containing 30 nM target DNA substrate with 30 nM IR DNA substrate and 25 nM transposase. Reactions were incubated for two hours at 30°C in the same buffer used for the cleavage assay. For target integration assays of dissolved crystals, reactions were set up as above but without IR DNA and transposase. Reactions were started by the addition of 2 μl of dissolved crystals diluted 1 in 10, 1 in 20 or 1 in 40 in 25 mM tris-HCl (pH 7.5), 0.25 M KCl and 50% glycerol (final protein concentration estimated as 60, 30, and 15 nM) and incubated for 2 hr at 30°C. The products were separated on an 8% denaturing polyacrylamide gel and fluorescein was visualized as above.

NTS Cleavage Assays with Protein Mixtures

The NTS cleavage dsDNA substrate was prepared by annealing a fluorescently labeled 100-mer NTS oligonucleotide with the sequence 5′-acgggctccagtcggtgaatatagaaactaTCAGGT GTA CAA GTA TGA AAT GTC GTT TTTTT TAAATCAAAAAACACG TTAAATTTTGTG GAAAAAGAAA plus a 6-carboxyfluorescein tag at the 5′ end with the complementary 100 nt TS.

Protein stocks were diluted in a buffer containing 12.5 mM Tris pH 7.5, 125 mM KCl and 50% glycerol to a concentration 20-fold higher than the final value (e.g., 600 nM stock for 30 nM final). Diluted proteins were mixed in equal volumes and incubated on ice for 10 min. Cleavage assays were initiated by adding 2 μl protein mixture to 20 μl reactions containing 30 nM dsDNA substrate as described above. Reactions were incubated for 15 min at 30°C and stopped by the addition of 20 μl of loading dye. The products were separated on an 8% denaturing polyacrylamide gel and the fluorescein label was detected as above. Reactions where the total protein concentration was kept at 30 nM were carried out as above except that 600 nM stocks of each protein were mixed together in different ratios and incubated on ice before addition to reactions.

EMSA

Reactions were set up exactly as above with the addition of 50 μg/ml poly dI-dC as competitor DNA. Reactions were incubated for 30 min at 30°C, placed on ice and then loaded directly onto native 6% polyacrylamide gels containing 10% glycerol in 0.5 x Tris-Acetate-EDTA (TAE) buffer. Gels were run at 200V for 3.5 hr at 4°C.

Sequence Alignments

Sequence alignments were carried out with T-coffee (http://www.ebi.ac.uk/t-coffee) followed by manual alignment based on secondary structure elements, either determined experimentally (for Mos1 and Tc3) or predicted (for SETMAR and Sleeping Beauty) using the PredictProtein server (http://cubic.bioc.columbia.edu/predictprotein).

ACCESSION NUMBERS

The Mg^{2+} and Mn^{2+} bound PEC structures are deposited in the Protein Data Bank with accession numbers 3HOS and 3HOT, respectively.

SUPPLEMENTAL DATA

Supplemental Data include eight figures and Supplemental References and can be found with this article online at http://www.cell.com/supplemental/S0092-8674(09)00851-4.

ACKNOWLEDGMENTS

This work was funded by the Wellcome Trust and the Medical Research Council. X-ray data were collected at the ESRF (beam-lines BM14 and ID23-1), and we thank Martin Walsh for his expert assistance. We are grateful to Marshall Stark, Jean Beggs, and Adrian Bird for their insightful comments on the manuscript.

Received: November 28, 2008
Revised: April 24, 2009
Accepted: July 2, 2009
Published: September 17, 2009

REFERENCES

Ason, B., and Reznikoff, W.S. (2002). Mutational analysis of the base flipping event found in Tn5 transposition. J. Biol. Chem. *277*, 11284–11291.

Auge-Gouillou, C., Brillet, B., Hamelin, M.H., and Bigot, Y. (2005). Assembly of the mariner Mos1 synaptic complex. Mol. Cell. Biol. *25*, 2861–2870.

Beese, L.S., and Steitz, T.A. (1991). Structural basis for the 3′-5′ exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. EMBO J. *10*, 25–33.

Biemont, C., and Vieira, C. (2006). Genetics: junk DNA as an evolutionary force. Nature *443*, 521–524.

Cordaux, R., Udit, S., Batzer, M.A., and Feschotte, C. (2006). Birth of a chimeric primate gene by capture of the transposase gene from a mobile element. Proc. Natl. Acad. Sci. USA *103*, 8101–8106.

Davies, D.R., Goryshin, I.Y., Reznikoff, W.S., and Rayment, I. (2000). Three-dimensional structure of the Tn5 synaptic complex transposition intermediate. Science *289*, 77–85.

Dawson, A., and Finnegan, D.J. (2003). Excision of the Drosophila mariner transposon Mos1. Comparison with bacterial transposition and V(D)J recombination. Mol. Cell *11*, 225–235.

Dupuy, A.J., Akagi, K., Largaespada, D.A., Copeland, N.G., and Jenkins, N.A. (2005). Mammalian mutagenesis using a highly mobile somatic Sleeping Beauty transposon system. Nature *436*, 221–226.

Grundy, G.J., Hesse, J.E., and Gellert, M. (2007). Requirements for DNA hairpin formation by RAG1/2. Proc. Natl. Acad. Sci. USA *104*, 3078–3083.

Guimond, N., Bideshi, D.K., Pinkerton, A.C., Atkinson, P.W., and O'Brochta, D.A. (2003). Patterns of Hermes transposition in Drosophila melanogaster. Mol. Genet. Genomics *268*, 779–790.

Ivics, Z., and Izsvak, Z. (2006). Transposons for gene therapy!. Curr. Gene Ther. *6*, 593–607.

Kapitonov, V.V., and Jurka, J. (2005). RAG1 core and V(D)J recombination signal sequences were derived from Transib transposons. PLoS Biol. *3*, e181. 10.1371/journal.pbio.0030181.

Lampe, D.J., Churchill, M.E., and Robertson, H.M. (1996). A purified mariner transposase is sufficient to mediate transposition in vitro. EMBO J. *15*, 5470–5479.

Lipkow, K., Buisine, N., Lampe, D.J., and Chalmers, R. (2004). Early intermediates of mariner transposition: catalysis without synapsis of the transposon ends suggests a novel architecture of the synaptic complex. Mol. Cell. Biol. *24*, 8301–8311.

Lohe, A.R., and Hartl, D.L. (1996). Autoregulation of mariner transposase activity by overproduction and dominant-negative complementation. Mol. Biol. Evol. *13*, 549–555.

Lovell, S., Goryshin, I.Y., Reznikoff, W.R., and Rayment, I. (2002). Two-metal active site binding of a Tn5 transposase synaptic complex. Nat. Struct. Biol. *9*, 278–281.

Luo, G., Ivics, Z., Izsvak, Z., and Bradley, A. (1998). Chromosomal transposition of a Tc1/mariner-like element in mouse embryonic stem cells. Proc. Natl. Acad. Sci. USA *95*, 10769–10773.

Luscombe, N.M., Laskowski, R.A., and Thornton, J.M. (2001). Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. Nucleic Acids Res. *29*, 2860–2874.

Mouw, K.W., Rowland, S.J., Gajjar, M.M., Boocock, M.R., Stark, W.M., and Rice, P.A. (2008). Architecture of a serine recombinase-DNA regulatory complex. Mol. Cell *30*, 145–155.

Naumann, T.A., and Reznikoff, W.S. (2000). Trans catalysis in Tn5 transposition. Proc. Natl. Acad. Sci. USA *97*, 8944–8949.

Plasterk, R.H., Izsvak, Z., and Ivics, Z. (1999). Resident aliens: the Tc1/mariner superfamily of transposable elements. Trends Genet. *15*, 326–332.

Rice, P.A., and Baker, T.A. (2001). Comparative architecture of transposase and integrase complexes. Nat. Struct. Biol. *8*, 302–307.

Richardson, J.M., Dawson, A., O'Hagan, N., Taylor, P., Finnegan, D.J., and Walkinshaw, M.D. (2006). Mechanism of Mos1 transposition: insights from structural analysis. EMBO J. *25*, 1324–1334.

Richardson, J.M., Finnegan, D.J., and Walkinshaw, M.D. (2007). Crystallization of a Mos1 transposase-inverted-repeat DNA complex: biochemical and preliminary crystallographic analyses. Acta Crystallograph. Sect. F. Struct. Biol. Cryst. Commun. *63*, 434–437.

Richardson, J.M., Zhang, L., Marcos, S., Finnegan, D.J., Harding, M.M., Taylor, P., and Walkinshaw, M.D. (2004). Expression, purification and preliminary crystallographic studies of a single-point mutant of Mos1 mariner transposase. Acta Crystallogr. D Biol. Crystallogr. *60*, 962–964.

Robert, V., and Bessereau, J.L. (2007). Targeted engineering of the Caenorhabditis elegans genome following Mos1-triggered chromosomal breaks. EMBO J. *26*, 170–183.

Robinson, A.S., Franz, G., and Atkinson, P.W. (2004). Insect transgenesis and its potential role in agriculture and human health. Insect Biochem. Mol. Biol. *34*, 113–120.

Roth, D.B., Menetski, J.P., Nakajima, P.B., Bosma, M.J., and Gellert, M. (1992). V(D)J recombination: broken DNA molecules with covalently sealed (hairpin) coding ends in scid mouse thymocytes. Cell *70*, 983–991.

Savilahti, H., and Mizuuchi, K. (1996). Mu transpositional recombination: donor DNA cleavage and strand transfer in trans by the Mu transposase. Cell *85*, 271–280.

Shevchenko, Y., Bouffard, G.G., Butterfield, Y.S., Blakesley, R.W., Hartley, J.L., Young, A.C., Marra, M.A., Jones, S.J., Touchman, J.W., and Green, E.D. (2002). Systematic sequencing of cDNA clones using the transposon Tn5. Nucleic Acids Res. *30*, 2469–2477.

Slotkin, R.K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. Nat. Rev. Genet. *8*, 272–285.

Swanson, P.C. (2001). The DDE motif in RAG-1 is contributed in trans to a single active site that catalyzes the nicking and transesterification steps of V(D)J recombination. Mol. Cell. Biol. *21*, 449–458.

van Luenen, H.G., Colloms, S.D., and Plasterk, R.H. (1994). The mechanism of transposition of Tc3 in C. elegans. Cell *79*, 293–301.

van Pouderoyen, G., Ketting, R.F., Perrakis, A., Plasterk, R.H., and Sixma, T.K. (1997). Crystal structure of the specific DNA-binding domain of Tc3 transposase of C.elegans in complex with transposon DNA. EMBO J. *16*, 6044–6054.

Watkins, S., van Pouderoyen, G., and Sixma, T.K. (2004). Structural analysis of the bipartite DNA-binding domain of Tc3 transposase bound to transposon DNA. Nucleic Acids Res. *32*, 4306–4312.

Xu, H.E., Rould, M.A., Xu, W., Epstein, J.A., Maas, R.L., and Pabo, C.O. (1999). Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. Genes Dev. *13*, 1263–1275.

Yang, W., Lee, J.Y., and Nowotny, M. (2006). Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity. Mol. Cell *22*, 5–13.

Zhou, L., Mitra, R., Atkinson, P.W., Hickman, A.B., Dyda, F., and Craig, N.L. (2004). Transposition of hAT elements links transposable elements and V(D)J recombination. Nature *432*, 995–1001.